unclassified

SECURITY CLASSIFICATION OF THIS PAGE

DTIC ELECTE JAN 09 1991 S B D

REPORT DOCUMENTATION PAGE

AD-A231 110

1a. REPORT SECURITY CLASSIFICATION	1b.
2a. SECURITY CLASSIFICATION AUTHORITY	3. Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 90 1220
6a. NAME OF PERFORMING ORGANIZATION Washington State University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) Pullman, Washington 99164-3113	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR; 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-88-0180
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
61102F	2304	A8	

11. TITLE (Include Security Classification)

Rapidly Convergent Algorithms for Nonsmooth Optimization

12. PERSONAL AUTHOR(S) Robert Mifflin

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 16/6/88 TO 30/9/90	90/1/12	5

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The research supported by this grant has continued the development of efficient methods for solving optimization problems involving implicitly defined functions that are not everywhere differentiable.

Progress has been made on extending a rapidly convergent algorithm for the single variable case to the n variable case. A specialization of this research has produced a new two matrix quasi-Newton method for smooth minimization.

Also, a new fast method has been developed for the single variable case where only function, and not subderivative, values are available.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
~~Maj. James M. Crowley~~ Neal Glassman	(202) 767-5026	NM

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

The research conducted under Grant number AFOSR-88-0180 during the period 16 June 1988 to 30 September 1990 is partially documented in [2], [3], [12], [13] and the remainder of this report. It is related to previous work in [1], [5], [6], [7], [8], [9], [10], [11] and [14].

Optimization problems with functions that are not everywhere differentiable often arise when certain techniques are applied to large or complicated nonlinear programming problems in order to convert them to a sequence of smaller or less complex problems. These transformation techniques include decomposition, nested dissection, relaxation, duality and/or exact L_1 penalty methods and often lead to functions to be minimized which are implicitly defined. Being able to solve such problems gives an analyst flexibility in modeling a problem for solution and the ability to exploit parallel processing in computation. Hence, it is important for practical applications to be able to solve such problems and it is the goal of the project to develop efficient solution methods.

Recent joint work with J.-J. Strodiot (Namur, Belgium) has produced practically useful and theoretically satisfying ideas for solving single variable minimization problems using function, but not derivative, values. This effort has produced, what is probably, the first instance of a function-value-only method for nonsmooth functions with proven rapid convergence. The corresponding paper [13] is being revised in order to append some figures to illustrate various cases mentioned in the introduction and considered in the proofs. The revision will be submitted for publication as soon as the retyping is completed. Our previous work on this subject which introduced a safeguarded bracketing technique and developed some quadratic approximation results appeared in [14].

The above-mentioned work stems from a bracketing method using generalized derivatives at the bracket end points developed and tested in [6], [8], [10] and [11]. Results from this work and [7] already appear in a textbook by K.G. Murty [15; pp. 410-421]. A new result [12] in this area is scheduled to appear in Vol. 49, No. 2 of *Mathematical Programming.* The result states that either the next iterate is superlinearly closer to the solution than both of the current bracket end points or the length of the next bracket is superlinearly shorter than that of the current bracket.

Recent research has shown us more specifically what needs to be looked at carefully in order to develop an algorithm having better than linear convergence for n-variable nonsmooth minimization problems. The problems we are interested in are those whose objectives have an underlying piecewise C^2 structure that is not explicitly known. The algorithm we are working on is a 2nd order bundle method

where each iterate depends on a bundle of previous iterates and their corresponding function and subgradient values. The 2nd order nature has to do with the method also employing a bundle of associated $n \times n$ Hessian matrix approximations. The main idea for determining an iterate is to solve a 1st order quadratic programming (QP) subproblem and then to modify the solution by a 2nd order correction resulting from the solution of one linear system whose data depends the constraint multiplier values from the QP problem and on the Hessian estimates associated with positive multipliers. A better than linear convergence result as in [9] based upon this new idea was presented in the Minisymposium on Nonsmooth Optimization at the April 1989 SIAM Conference on Numerical Optimization held in Boston. One difficulty with this approach is that the 2nd order system matrix may not be positive definite.

To study this issue the principal investigator and his Ph.D. student research assistant have worked out a method for smooth (one C^2 piece) minimization which employs both BFGS and symmetric rank one (SR1) updated Hessian estimates. The BFGS matrix is initialized and updated to be positive definite, whereas the SR1 matrix is initialized to be zero and may not be positive definite at other iterations. Powell [16] has shown that in some situations the BFGS method can be slow to approximate a Hessian accurately, because this accuracy depends upon the generation of conjugate directions which in turn depends on exact line searches. The accuracy of the SR1 method only depends upon using linearly independent search directions and, hence, the SR1 matrix is used in the second order model of the function in our method. To solve the resulting linear system with a possibly indefinite matrix, a truncated preconditioned conjugate gradient [18] method is employed. The preconditioner matrix is the positive definite BFGS matrix. We can show that if this preconditioned conjugate gradient (PCG) submethod is initialized with the zero vector then its 1st subdirection is the ordinary BFGS direction and every subiterate and subdirection generated is a descent direction for the objective function at its current point. Hence, the PCG submethod can be terminated at any time with a descent direction which is a modification of the BFGS direction designed to reduce the value of the SR1 2nd order model. The PCG submethod must be terminated if a subdirection is generated which is a direction of nonpositive curvature with respect to the SR1 matrix. Using such a search direction at a point in a region where the objective function is nonconvex may be beneficial. Near the solution of the main problem one can expect the BFGS and SR1 matrices to be similar so the PCG algorithm need not perform many subiterations.

For an initial test of the above ideas we used the following three-variable version of a two-variable function due to Byrd, Nocedal and Yuan[4]:

$$f(x) = \frac{1}{2}x^T x + \sigma(\frac{1}{2}x^T A x)^2$$

where

$$A = \begin{bmatrix} 5. & 1. & 0 \\ 1. & 3. & 1. \\ 0 & 1. & 5. \end{bmatrix} \text{ and } \sigma = 0.1.$$

The starting point was $x_0 = (\cos 70°, \sin 70°, 1.0)$. The BFGS matrix was initialized as a diagonal matrix with diagonal elements $(1., 10^4, 10.)$ and the SR1 was initialized as the zero matrix in order to have a neutral start.

We modified the code UNCMIN due to Schnabel, Koontz and Weiss[17] by appending subroutines to update the $SR1$ matrix and perform the preconditioned conjugate gradient iteration. We chose the option in UNCMIN which updates factors of the BFGS matrix. Also, we replaced the line search subroutine by one written by Moré and Thuente which was changed slightly to impose only the Wolfe [19] stopping conditions.

A user-defined parameter l was included to define the maximum number of conjugate gradient subiterations allowed at each major iteration. When $l = 0$, our algorithm becomes a pure BFGS method.

The four runs summarized in the following table had the same first point x_1, as well as initial point x_0, and were terminated with iterate x_k when

$$\max_{1 \le j \le 3} \left| \frac{\partial f(x_k)}{\partial x_j} \right| \le 10^{-9}.$$

	$l = 0$ (BFGS)	$l = 1$	$l = 2$	$l = 3$
iterations	32	19	14	14
evaluations	34	23	17	20
SR1-error	3×10^{-7}	1×10^{-7}	6×10^{-5}	6×10^{-6}
BFGS-error	6×10^{-2}	2×10^{-3}	6×10^{-2}	6×10^{-2}

The first row gives the number of iterations, the second gives the number of evaluations of f and ∇f, and the third and fourth give the terminal values of

$$\max_{i,j} |(S - I)_{ij}| \text{ and } \max_{i,j} |(B - I)_{ij}|,$$

respectively, where the identity matrix I equals $\nabla^2 f$ at the solution $(0, 0, 0)$, S is the final SR1 matrix and B is the final BFGS matrix. The difference in the BFGS and SR1 errors should be noted along with the overall ability of the algorithm with

$l > 0$ to exploit the better approximation of the SR1 matrix. The above results were presented at the May 1990 TIMS/ORSA Joint National Meeting held in Las Vegas and in a slightly generalized form at the November 1990 Pacific West Optimization Seminar in Seattle.

Finally, for solving the quadratic programming subproblem in the nonsmooth algorithm it may be beneficial to develop a special purpose method based on the Ph.D. dissertation of A. Al-Saket [1]. Two general papers [2] and [3] on this subject have been prepared for publication submission.

REFERENCES

[1] A. Al-Saket, An algorithm with degeneracy resolution for solving certain quadratic programming problems, Dissertation, Washington State University, Pullman, Washington, 1985.

[2] A. Al-Saket and R. Mifflin, An algorithm for solving certain quadratic programming problems, Dept. of Mathematics, University of Jordan, Amman, Jordan, 1989.

[3] A. Al-Saket and R. Mifflin, An algorithm with degeneracy resolution for solving certain quadratic programming problems, Dept. of Mathematics, University of Jordan, Amman, Jordan, 1989.

[4] R.H. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Anal.* **24**, (1987), 1171-1190.

[5] N. Gupta, A higher than first order algorithm for nonsmooth constrained optimization, Dissertation, Washington State University, Pullman, Washington, 1985.

[6] C. Lemarechal and R. Mifflin, A globally and superlinearly convergent algorithm for one-dimensional mimimization of convex functions, *Mathematical Programming* **24**, (1982), 241-256.

[7] R. Mifflin, A superlinearly convergent algorithm for one-dimensional constrained minimization problems with convex functions, *Mathematics of Operations Research* **8**, (1983), 185-195.

[8] R. Mifflin, Stationarity and superlinear convergence of an algorithm for univariate locally Lipschitz constrained minimization, *Mathematical Programming* **28**, (1984), 50-71.

[9] R. Mifflin, Better than linear convergence and safeguarding in nonsmooth minimization, in P. Thoft-Christensen, ed., **System Modeling and Optimization** (Springer-Verlag, New York, 1984).

[10] R. Mifflin, The solution of a nested nonsmooth optimization problem, **Nondifferentiable Optimization: Motivation and Applications**, V. Demyanov and D. Pallasche eds , Springer-Verlag, New York, 1985.

[11] R. Mifflin, An implementation of an algorithm for univariate minimization and an application to nested optimization, *Mathematical Programming Study* **31** (1987), 155-165.

[12] R. Mifflin, On superlinear convergence in univariate nonsmooth minimization, to appear in *Mathematical Programming.*

[13] R. Mifflin and J.-J. Strodiot, A rapidly convergent five point algorithm for univariate minimization, Report of the Department of Mathematics, Facultes Universitaires de Namur, Namur, 1986.

[14] R. Mifflin and J.-J. Strodiot, A safeguarded bracketing technique for obtaining convergence and preserving rapid convergence in univariate minimization, *Mathematical Programming* **43** (1989), 117-130.

[15] K.G. Murty, **Linear Complementarity, Linear and Nonlinear Programming,** Heldermann Verlag, Berlin, 1988.

[16] M.J.D. Powell, How bad are the BFGS and DFP methods when the objective function is quadratic? *Mathematical Programming* **34** (1986), 34-47.

[17] R. B. Schnabel, B.F. Weiss, and J.E. Koontz, A modular system of algorithms for unconstrained minimization, University of Colorado Computer Science, TR CU-CS-240-82, 1982.

[18] T. Steihaug, The conjugate gradient method and trust regions in large scale optimization, *SIAM J. Numer. Anal.* **20** (1983), 626-637.

[19] P. Wolfe, Convergence conditions for ascent methods, SIAM Review II (1969), 226-235.